Milk intake and incident stroke and coronary heart disease in populations of European descent: A Mendelian Randomization study.

L.E.T. Vissers[1], I. Sluijs[1], S. Burgess[2,3], N.G. Forouhi[4], H. Freisling[5], F. Imamura[4], T.K. Nilsson[6], F. Renström[7,8], E. Weiderpass[5], K. Aleksandrova[11], C.C. Dahm[12], A. Perez-Cornago[13], M.B. Schulze[9,10,11], T.Y.N. Tong[13], D. Aune[14,15,16], C. Bonet[17], J.M.A. Boer[18], H. Boeing[9], M.D. Chirlaque[19,20], M.I. Conchi[21,22], L. Imaz[23,24], S. Jäger[9,10], V. Krogh[25], C. Kyrø[26], G. Masala[27], O. Melander[28], K. Overvad[12,29], S. Panico[30], M.J. Sánches[20,31,32,33], E. Sonestedt[28], A. Tjønneland[26], I. Tzoulaki[14], W.M.M. Verschuren[1,18], E. Riboli[34], N.J. Wareham[4], J. Danesh[4], A.S. Butterworth[2,35], Y.T. van der Schouw[1]

[1]Julius Centre for Health Sciences and Primary Care, University Medical Centre Utrecht, University Utrecht, Utrecht, the Netherlands

[2]Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK

[3]MRC Biostatistics Unit, University of Cambridge, Cambridge, UK

[4]MRC Epidemiology Unit, University of Cambridge, Cambridge, UK

[5]International Agency for Research on Cancer, Lyon, France

[6]Department of Medical Biosciences/Clinical Chemistry, Umeå university, Umeå, Sweden

[7]Department of Biobank Research, Umeå University, Umeå, Sweden

[8]Division of Endocrinology and Diabetes, Cantonal Hospital St. Gallen, St. Gallen, Switzerland

[9]German Institute of Human Nutrition Potsdam-Rehbruecke, Nuthetal, Germany

[10]German Center for Diabetes Research (DZD), Neuherberg, Germany

This peer-reviewed article has been accepted for publication but not yet copyedited or typeset, and so may be subject to change during the production process. The article is considered published and may be cited using its DOI

10.1017/S0007114521004244

The British Journal of Nutrition is published by Cambridge University Press on behalf of The Nutrition Society

[11]Germany Institute of Nutritional Sciences, University of Potsdam, Nuthetal, Germany

[12]Department of Public Health, Aarhus University, Aarhus, Denmark

[13]Cancer Epidemiology Unit, Nuffield Department of Population Health, University of Oxford, Oxford, UK

[14]Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, London, UK

[15]Department of Nutrition, Bjørknes University College, Oslo, Norway

[16]Department of Endocrinology, Morbid Obesity and Preventive Medicine, Oslo University Hospital, Oslo, Norway

[17]Unit of Nutrition and Cancer. Cancer Epidemiology Research Program, Catalan Institute of Oncology-IDIBELL, L'Hospitalet de Llobregat, Barcelona, Spain.

[18]National Institute for Public Health and the Environment, Bilthoven, The Netherlands

[19]Department of Epidemiology, Regional Health Council, IMIB-Arrixaca, Murcia University, Murcia, Spain.

[20]CIBER in Epidemiology and Public Health (CIBERESP), Madrid, Spain.

[21]Navarra Public Health Institute - IdiSNA, Pamplona, Spain

[22]Research Network on Health Services in Chronic Diseases (REDISSEC), Pamplona, Spain

[23]Ministry of Health of the Basque Government, Public Health Division of Gipuzkoa, Donostia-San Sebastian, Spain

[24]Biodonostia Health Research Institute, Donostia-San Sebastian, Spain

[25]Epidemiology and prevention Unit, Fondazione IRCCS Instituto Nazionale dei Tumori, Milano, Italy

[26]Danish Cancer Society Research Center, Copenhagen, Denmark.

[27]Cancer Risk Factors and Life-Style Epidemiology Unit, Institute for Cancer Research, Prevention and Clinical Network - ISPRO, Florence, Italy

[28]Lund University, Department of Clinical Sciences, Malmö, Sweden

[29]Department of Cardiology, Aalborg University Hospital, Aalborg, Denmark

[30]Dipartemento di medicina clinica e chirurgia, Federico II University, Naples, Italy

[31]Andalusian School of Public Health (EASP), Granada, Spain

[32]Instituto de Investigación Biosanitaria de Granada, Granada, Spain

[33]Universidad de Granada, Granada, Spain

[34]School of Public Health, Imperial College London, UK

[35]Department of Medical Statistics, London School of Hygiene and Tropical Medicine, London, UK.

Corresponding author: prof. Yvonne van der Schouw, PhD, Julius Centre for Health Sciences and Primary Care, University Medical Centre Utrecht, STRT 6.131, PO Box 85500, 3508GA Utrecht, The Netherlands, T: 0031 88 7568181, E: Y.T.vanderSchouw@umcutrecht.nl

The data underlying this article were provided by the EPIC-CVD consortium. Data will be shared on request to the corresponding author with permission of the EPIC-CVD consortium.

Abstract:

Higher milk intake has been associated with a lower stroke risk, but not with risk of coronary heart disease (CHD). Residual confounding or reverse causation cannot be excluded. Therefore, we estimated the causal association of milk consumption with stroke and CHD risk through instrumental variable (IV) and gene-outcome analyses. IV analysis included 29,328 participants (4,611 stroke; 9,828 CHD) of the EPIC-CVD (8 European countries) and EPIC-NL case-cohort studies. rs4988235, a lactase persistence (LP) single nucleotide polymorphism which enables digestion of lactose in adulthood was used as genetic instrument. Intake of milk was first regressed on rs4988235 in a linear regression model. Next, associations of genetically predicted milk consumption with stroke and CHD were estimated using Prentice-weighted Cox regression. Gene-outcome analysis included 777,024 participants (50,804 cases) from MEGASTROKE (including EPIC-CVD), UK Biobank and EPIC-NL for stroke, and 483,966 participants (61,612 cases) from CARDIoGRAM, UK Biobank and EPIC-CVD and EPIC-NL for CHD. In IV analyses, each additional LP allele was associated with a higher intake of milk in EPIC-CVD (β=13.7 g/day; 95%CI: 8.4-19.1) and EPIC-NL (36.8 g/day; 20.0-53.5). Genetically predicted milk intake was not associated with stroke (HR per 25 g/day 1.05; 95%CI: 0.94-1.16) or CHD (1.02; 0.96-1.08). In gene-outcome analyses, there was no association of rs4988235 with risk of stroke (odds ratios 1.02; 0.99-1.05) or CHD (0.99; 0.95-1.03). Current Mendelian Randomization analysis does not provide evidence for a causal inverse relationship between milk consumption and stroke or CHD risk.

Keywords: Milk, dairy, CHD, stroke, Mendelian Randomization

Introduction

Higher intake of milk has been associated with a modestly lower risk of incident stroke, but not coronary heart disease (CHD), in a meta-analysis of observational studies[1] and more recently in EPIC-CVD[2]. However, potential confounding and reverse causation cannot be excluded[3] and therefore any preventive effect of milk consumption in the development of stroke remains debatable.

A Mendelian Randomization (MR) approach[3] could be used to elucidate the causality of this association. An MR study is essentially an instrumental variable (IV) analysis, using one or more genetic variants as IV. The assumptions in an MR study are as follows: 1) The genetic variant is associated with the determinant of interest. 2) The genetic variant is not associated with known or unknown confounders of the determinant-outcome relationship. 3) There is no pathway from the genetic variant to disease that does not include the exposure of interest.

A genetic variant (rs4988235, -13910C>T) near the lactase gene fulfils the first IV assumption, as this variant has been associated in European populations with adult lactase persistence (LP)[4, 5], i.e. the continued capacity to produce lactase in the intestinal lumen in adulthood. In absence of this enzyme, people cannot break down lactose from dairy products, leading to gastro-intestinal symptoms when they consume milk. Alleles conferring LP have been associated with a higher intake of milk in most European-ancestry cohorts[6-12]. Furthermore, our previous MR analysis investigating diabetes risk in EPIC-InterAct did not show violation of the second IV assumption[10].

Previous MR studies in people of European descent reported no association between LP-associated milk intake and CHD[11, 13] or with total cardiovascular disease (CVD)[14]. A potential causal association between milk consumption and stroke risk has not yet been tested using MR. We therefore used rs4988235 in an IV analysis to investigate whether there is a causal relationship between LP-predicted milk intake and risk of total and ischaemic stroke and CHD in studies of participants of European descent. In addition, we performed gene-outcome analyses for the association of rs4988235 with stroke and CHD using data from the EPIC studies, supplemented with data from large-scale genetic consortia lacking quantitative data on habitual milk intake.

Methods

We performed a one-sample MR analysis, further described as IV analysis, in studies for which we had access to individual participant phenotype information for habitual intake of milk and other foods, on presence of cardiovascular risk factors and stroke and CHD outcomes, as well information on rs4988235. We performed gene-outcome analysis using data from several studies lacking quantitative data on habitual milk intake. The different studies are further described below.

Data for IV analysis

We used sub-populations of the EPIC study: the EPIC-CVD case-cohort study and the Dutch EPIC-NL study.

EPIC-CVD: Among participants with a stored blood sample available, a representative subcohort was selected[15, 16]. During follow-up of the EPIC study, incident stroke and CHD cases occurred and these participants were added to the EPIC-CVD case-cohort population[17]. We included participants with information on intake of milk ((semi-)skimmed or full-fat, regardless of fermentation) from region-specific diet questionnaires, the development and validity of which were reported previously[15, 18, 19] and on incident stroke and CHD among those who had data on rs4988235. The EPIC-CVD case-cohort study population for this analysis consists of 13,114 subcohort participants including 397 stroke cases and 521 CHD cases from 8 European countries, plus additional cases of stroke (n=3,737) and CHD (n=8,985) (Supplemental methods, Figure 1.) Of the 4,134 total stroke cases in the full case-cohort population, 2,746 (66%) were identified as ischaemic strokes. In EPIC-CVD, variant rs4988235 was genotyped using the Illumina HumanCore Exome Chip array (n=16,685), or genotype dosage was imputed for those who were genotyped on the Illumina HumanCoreExome, Illumina OmniExomeExpress, or Illumina Quad660 array (n=8,055, impute info~0.42).

EPIC-NL: In the Dutch EPIC cohort[20], we selected an additional random subcohort. Incident CHD and stroke cases during follow-up of the EPIC-NL study that occurred after the follow-up for EPIC-CVD was completed were added to obtain the EPIC-NL case-cohort. Data from 3,331 participants with information on milk consumption ((semi-)skimmed or full-fat, regardless of fermentation), CHD and stroke incidence and imputed rs4988235 genotype was

used for analysis, including 843 CHD and 567 total stroke cases, of which 410 (72%) were identified as ischaemic strokes. We genotyped participants using the Illumina Global Screening Assay BeadArray and rs4988235 was imputed at high quality (impute info=0.91).

The EPIC-CVD study cohort and additional participants from EPIC-NL are described in more detail in the Supplemental Methods. The EPIC studies were approved by institutional ethical review boards. All participants gave written informed consent prior to inclusion.

Data for gene-outcome analysis

For the gene-outcome analysis, we included publically available data from various consortia, namely UK Biobank, CARDIoGRAM GWAS and MEGASTROKE. We additionally included gene-outcome associations from EPIC-CVD and EPIC-NL after exclusion of the population overlap with CARDIoGRAM and MEGASTROKE (Supplemental Table 1). A description of these cohorts can be found in the supplemental methods.

Data analysis

In descriptive analyses, we assessed baseline characteristics and dietary intakes in the EPIC-CVD and EPIC-NL subcohort. We also tested for deviation from Hardy Weinberg Equilibrium (HWE)[21] for rs4988235.

We investigated IV assumptions[3], namely whether rs4988235 was reliably associated with milk consumption and whether it was associated to cardiometabolic risk factors. For continuous variables, we reported a β and 95% confidence interval and for dichotomous variables, we reported an odds ratio and 95% confidence interval. We conducted linear regression analyses stratified by imputed and hard-call data and pooled the outcomes with fixed-effects meta-analysis within EPIC-CVD. The models were adjusted for sex, age, the first two genetic principal components (PC)[22], and study centre, assuming an additive effect of rs4988235[23]. Additional information can be found in the Supplemental Methods. Dairy non-consumers were not excluded from analysis since avoidance of dairy could be influenced by rs4988235 genotype. None of the aforementioned variables were missing for any of the participants.

Next, we performed a two-stage least squared (2SLS) IV analysis[24], to investigate causality of the association of LP- predicted milk consumption with risk of stroke and CHD

separately. LP predicted milk consumption was calculated for each participant by using rs4988235, genotyping platform, sex, age, the first two genetic PC and study centre as predictors in a linear regression model. LP-predicted milk consumption was scaled to obtain interpretable estimates of the hazard ratio for CHD and stroke per 25 g/day increase in milk consumption. We fitted a Prentice-weighted Cox proportional hazard model to take the case-cohort design into account[25] with age as underlying time scale and adjusted for sex, the first two genetic PC and study centre. We tested the effect of including additional genetic PC in our model, but this reduced explained variance (R squared statistic) of the model and did not affect effect estimates. Analyses were performed by country in EPIC-CVD. Within countries, the analysis was stratified according to hard call versus imputed data for rs4988235 and subsequently pooled across strata using fixed effects. Country-specific results from EPIC-CVD and the additional EPIC-NL results were then pooled with inverse variance weights in a random effect meta-analysis using restricted maximum likelihood estimation. The I squared (I^2) statistic was used to assess heterogeneity.

We repeated the aforementioned IV analysis under the assumption of dominance of LP[26], considering only participants with a C/C genotype lactase non persistent. We also performed a sensitivity analysis restricting our CHD outcomes to myocardial infarction. Since rs4988235 was associated with smoking in EPIC-NL, and smoking is unlikely to mediate an effect of milk consumption on CVD, we performed a post-hoc analysis, adjusting our MR estimate for smoking status (never versus ever smoker).

For the gene-outcome analysis regarding stroke, we used log odds ratios and standard errors for the association of rs4988235 with stroke from MEGASTROKE (which already included the EPIC-CVD data), and from UK Biobank (adjusted for the first 10 PC). In addition, the gene-outcome association between rs4988235 and stroke was investigated in EPIC-NL using a Prentice-weighted Cox regression with age as underlying time scale and identical covariates as in our IV analysis. The beta and standard errors from MEGASTROKE, UK Biobank and EPIC-NL were pooled with inverse variance weights in a random effect meta-analysis using restricted maximum likelihood estimation to obtain a final estimate for the association between rs4988235 and risk of stroke.

For the gene-outcome analysis regarding CHD, we used log odds and standard errors for the association of rs4988235 with CHD from CARDIoGRAM and UK Biobank (adjusted for first 10 PC). In addition, the gene-outcome association between rs4988235 and CHD was

investigated in EPIC-CVD and EPIC-NL using a Prentice-weighted Cox regression with the same approach and adjustment for covariates as in our IV analysis. We excluded EPIC-Norfolk from the EPIC-CVD analysis since this cohort was included in CARDIoGRAM.

The beta and standard errors from CARDIOGRAM, UK Biobank, and the combined EPIC studies were pooled with inverse variance weights in a random effect meta-analysis using restricted maximum likelihood estimation to obtain a relative risk for the association between rs4988235 and risk of CHD.

All analyses and data visualizations were performed in R[27] version 3.4.1, using the R packages survival[28], metafor[29], and forestplot[30].

Results

Descriptive analyses

The EPIC-CVD subcohort participants were on average 52 ± 9 years old and consisted of 60.7% women (Supplemental Table 2). Lactase persistence (rs4988235 C/T or T/T) ranged from 30.1% in Italy to 95.1% in Denmark. We observed deviation from HWE (at $p<0.05$) in the total EPIC-CVD subcohort, and in Denmark and Sweden (Supplemental Table 3). The EPIC-NL subcohort had an average age of 51 ± 11 years and consisted of 80.1% women (Supplemental Table 4). Lactase persistence occurred at a frequency of 90.3% and genotypes did not deviate from HWE ($p=0.20$).

Median milk intake was 160 g/day (interquartile range IQR; 39-295) in EPIC-CVD (Table 1), ranging from 32 (IQR 2-97) in Germany to 294 (IQR 149-440) in the UK (Supplemental Table 5), and 219 g/day (IQR 74-410) in EPIC-NL (Table 2). Dietary intake stratified by LP genotype is reported in Supplemental Tables 6 and 7.

Checking IV assumptions

Each additional T allele was associated with higher milk intake (EPIC-CVD: β 13.7 g/day (95%CI 8.4, 19.1); EPIC-NL: β 36.7 g/day (95%CI 19.8, 53.6)), but not with intake of total energy or with dairy products other than milk. Each T allele was also associated with lower meat intake (EPIC-CVD: β -7.3 g/day (95%CI -12.1, -2.5); EPIC-NL: β -4.4 g/day (95%CI -8.0, -0.9)) (Table 1 and 2).

In EPIC-CVD (Supplemental Table 8), each T allele was associated to a higher BMI (β 0.15 kg/m^2; 95%CI 0.04,0.27) and waist-to-hip ratio (β 2.4*10^{-3}; 95%CI 7.4*10^{-4}, 4.5*10^{-3}), and lower HDL cholesterol (β -0.02 mmol/L; 95%CI -0.03, -0.01) and lower odds of having a history of hypercholesterolemia (OR 0.90; 95%CI 0.83, 0.99). In EPIC-NL (Supplemental Table 9), rs4988235 was associated with higher odds of being a never smoker (OR 1.19, 95%CI 1.03-1.37) (Supplemental Table 9).

Stroke risk

In additive IV analysis, genetically predicted milk intake was not associated with risk of total stroke ($HR_{per\ 25\ g/day}$ 1.04, 95%CI: 0.94-1.16, I^2=23%) (Figure 1) or ischaemic stroke ($HR_{per\ 25\ g/day}$ 1.06, 95%CI: 0.93-1.21, I^2=22%) (Supplemental Figure 1), although confidence intervals are wide. We repeated analysis for total stroke under the assumption of a dominant effect of LP and results did not differ (Supplemental Figure 2). Additional adjustment for smoking status in the IV analysis of stroke risk in EPIC-NL yielded an HR of 1.15, 95%CI: 1.02-1.29 as compared to an HR of 1.14, 95%CI: 1.02-1.27 in original analysis. Gene-outcome analysis did not provide evidence for an association between rs4988235 and total stroke risk (OR 1.02, 95%CI: 0.99-1.05, I^2=39%) (Figure 2).

CHD risk

Genetically predicted milk intake was not associated with risk of CHD, with a pooled $HR_{per\ 25\ g/day}$ of 1.02 (95%CI: 0.96-1.08, I^2=0%) (Figure 3). No association with risk of myocardial infarction was found either with ($HR_{per\ 25\ g/day}$ 1.00, 95%CI: 0.93-1.07, I^2=0%) Results were not materially different when assuming a dominant effect of rs4988235 (Supplemental Figure 3). Additional adjustment for smoking status in the IV analysis of CHD risk in EPIC-NL yielded an HR of 1.00, 95%CI: 0.91-1.11 as compared to an HR of 1.00, 95%CI: 0.91-1.10 in original analysis). In gene-outcome analysis, we found no association of rs4988235 with CHD risk, with a pooled OR of 0.99 (95%CI: 0.95-1.04, I^2=73%) (Figure 4).

Discussion

In this IV and gene-outcome analysis among people of European descent, we did not find evidence for a causal relationship between milk consumption and risk of stroke or CHD.

Our study has several strengths. First, we used population-based cohorts with extensive phenotyping for our IV analysis. This allowed us to investigate the association between LP and intake of various (dairy) foods and cardiovascular risk factors, so we can attribute the observed association to a well-defined exposure[31]. In addition, we included a large number of participants in GWAS and population studies to investigate the association between rs4988235 and risk of stroke and CHD.

There are also limitations to address. First, previous studies have observed an association between genetic variation in the lactase region and consumption of total dairy[12, 14]. One of these studies showed a 30.3 g/day (95%CI: 21.3-29.3) higher total dairy intake per LP allele, and a 26.4 g/day (95%CI: 16.7-36.2) higher milk consumption, among 20,028 US participants. In EPIC-CVD, each additional LP allele was associated with a 13.6 g/day (95%CI: 8.4-18.8) higher milk intake on average, but not with dairy products other than milk (β 1.9 g/day (95%CI: -0.9-4.7). Milk likely explains most of the association with consumption of total dairy as milk has the highest lactose content[32]. However, there may be differences in the association of LP with dairy products other than milk between populations and his could affect findings from our gene-outcome analysis, since not all dairy products are the same in macro- and micronutrient content[33].

Second, we used dietary questionnaires to estimate milk intake, which are prone to both non-differential and differential measurement error. People with clinical symptoms of lactose intolerance may be more precise in their recall of dairy consumption, leading to differential misclassification. However, we expect misclassification to be non-differential between participants with and without later CVD. It is also likely that milk consumption in an individual is time-varying, but this is less relevant for an MR study since direction of the population effect is expected to remain the same. Also, we observed modest associations of rs4988235 with dietary and cardiovascular risk factors in EPIC-NL (e.g. more smoking) and EPIC-CVD (e.g. less meat intake), but we consider it unlikely that these modest associations reflect violation of the second IV assumption, as discussed in more detail in our previous EPIC-InterAct analysis[10]. Third, imputation quality for rs4988235 was low in a subset of the EPIC-CVD population

(n=8,055, impute info~0.42), so there may be LP genotype misclassification. Also, we observed deviation from Hardy-Weinberg equilibrium within EPIC-CVD, which could reflect selection bias is genetic population cohorts[21]. However, this deviation can be explained by the widely varying allele frequencies for rs4988235 between different European ancestry populations, and we have adjusted for potential resulting population stratification via principal components[22]. Lastly, outcome definitions for stroke and CHD were somewhat different between the cohorts used, but IV sensitivity analyses after restricting to ischemic stroke and myocardial infarction did not alter conclusions. Due to aforementioned limitations of our study we cannot exclude a small effect of milk consumption on risk of stroke or CHD.

Despite a modest association found between LP and higher WHR and BMI, which could be due to higher energy intake in LP persons[34], we did not observe an association with genetically predicted milk consumption on CHD risk. This lack of association is in line with findings from a meta-analysis of prospective cohort studies[1], a recent EPIC analysis[35], and a previous MR study among a Danish population[11].

Regarding total and ischaemic stroke, we did not find support for a protective effect of milk consumption in the IV or gene-outcome analysis, although confidence intervals for the IV analysis were wide. The biological mechanism of a protective effect of milk intake on stroke risk is thought to work via hypertension[36], potentially through intake of minerals such as calcium[37] and potassium[38]. We observed no association of lactase persistence with hypertension, or with systolic or diastolic blood pressure. In line with our findings, a previous MR study found no association between genetically predicted dairy consumption and hypertension[12]. However, a recent meta-analysis of 15 prospective cohort studies (4,381,604 participants, 25,377 stroke cases) reported an inverse association ($RR_{per\ 200\ gram/day}$ 0.92 , 95%CI 0.88-0.97) between consumption of milk and risk of stroke[1]. This could be because aforementioned association was mainly driven by East Asian populations, whereas a null association was observed in western populations. It could be hypothesized that people from East Asian descent benefit more from milk consumption due to genetic differences with people from European descent, or that increasing milk intake on top of a western diet – characterized by a high consumption of refined cereals, sugars and vegetable oils, and by consumption of dairy products, fatty meats and salt[39] – does not have a beneficial effect on stroke risk. Another hypothesis could be that the association between milk intake and risk of stroke is non-linear, although this cannot fully explain our results

since the meta-analysis suggests some protective effect of milk intake on stroke risk at any intake level[1]. We were unable to test the hypothesis of a non-linear association due to insufficient power to perform non-linear MR.

In conclusion, in IV analyses including 4,611 total stroke and 9,828 CHD cases and gene-outcome analyses including 50,804 stroke and 61,612 CHD cases, we did not find evidence of a causal relation between milk intake and risk of stroke or CHD in European populations. This suggests that the inverse association between milk intake and stroke in observational studies may be due to confounding. Given the stronger observational evidence for the association between milk intake and stroke in East Asian populations with generally lower milk intakes, future studies could focus on investigating the relationship between milk intake and stroke in people of East Asian descent, or on a potential non-linear association between milk intake and stroke.

Acknowledgements

We thank all EPIC participants and staff for their contribution to the study. We also thank staff from the EPIC-CVD coordinating centres for sample preparation and data handling.

This research has been conducted using the UK Biobank Resource (application number 29916).

Data on coronary artery disease have been contributed by CARDIoGRAMplusC4D investigators and have been downloaded from www.CARDIOGRAMPLUSC4D.ORG.

Financial Support

I Sluijs was supported by a personal Dr. Dekker postdoctoral grant (2015T019) from the Netherlands Heart Foundation. NGF and FI acknowledge core Medical Research Council Epidemiology Unit support (MC_UU_12015/5) and NGF acknowledges NIHR Biomedical Research Centre Cambridge: Nutrition, Diet, and Lifestyle Research Theme (IS-BRC-1215-20014).

The InterAct project was funded by the EU FP6 programme (grant number LSHM_CT_2006_037197), and provided the biomarker data in the subcohort that was used in the current study. These analyses were supported by Cancer Research UK (C8221/A19170). The coordination of EPIC is financially supported by the European Commission (DG-SANCO) and the International Agency for Research on Cancer. The national cohorts are supported by Danish

Cancer Society (Denmark); German Cancer Aid, German Cancer Research Center (DKFZ), Federal Ministry of Education and Research (BMBF), Deutsche Krebshilfe, Deutsches Krebsforschungszentrum and Federal Ministry of Education and Research (Germany); Associazione Italiana per la Ricerca sul Cancro-AIRC-Italy and National Research Council (Italy); Dutch Ministry of Public Health, Welfare and Sports (VWS), Netherlands Cancer Registry (NKR), LK Research Funds, Dutch Prevention Funds, Dutch ZON (Zorg Onderzoek Nederland), World Cancer Research Fund (WCRF), Statistics Netherlands (The Netherlands); Health Research Fund (FIS) PI13/00061 (EPIC-Granada) and, PI13/01162 (EPIC-Murcia), Regional Governments of Andalucía, Asturias, Basque Country, Murcia and Navarra, ISCIII Health Research Funds RD12/0036/0018 (cofounded by FEDER funds/European Regional Development Fund ERDF) (Spain); Swedish Cancer Society, Swedish Research Council and County Councils of Skåne and Västerbotten (Sweden); Cancer Research UK (14136 to EPIC-Norfolk; C570/A16491 and C8221/A19170 for EPIC-Oxford), Medical Research Council (1000143 to EPIC-Norfolk, MR/M012190/1 to EPIC-Oxford) (UK).

EPIC-CVD has been supported by the European Commission Framework Programme 7 (HEALTH-F2-2012-279233), the European Research Council (268834), the UK Medical Research Council (MR/L003120/1), the British Heart Foundation (RG13/13/30194 and RG/18/13/33946) and the National Institute for Health Research [Cambridge Biomedical Research Centre at the Cambridge University Hospitals NHS Foundation Trust].

The MEGASTROKE project received funding from sources specified at http://www.megastroke.org/acknowledgments.html.

Conflicts of Interest

Authors have no conflict of interest to declare

Authors Contributions

L.E.T.V. analyzed data and drafted the manuscript. L.E.T.V., I.S. and Y.T.vdS. had access to all data for this study. L.E.T.V., I.S., Y.T.vdS., S.B., N.G.F., H.F., F.I., T.K.N., F.R., E.W., K.A., C.D., A.P.C., M.B.S., T.Y.N.T, A.S.B. contributed to study conception, design and interpretation of data. All authors contributed to critical revision of the manuscript and approval of version to be published.

Literature Cited.

1. Soedamah-Muthu SS, de Goede J. Dairy Consumption and Cardiometabolic Diseases: Systematic Review and Updated Meta-Analyses of Prospective Cohort Studies. Curr Nutr Rep 2018;**7**(4):171-182.

2. Tong TYN, Appleby PN, Key TJ, Dahm CC, Overvad K, Olsen A, Tjonneland A, Katzke V, Kuhn T, Boeing H, Karakatsani A, Peppa E, Trichopoulou A, Weiderpass E, Masala G, Grioni S, Panico S, Tumino R, Boer JMA, Verschuren WMM, Quiros JR, Agudo A, Rodriguez-Barranco M, Imaz L, Chirlaque MD, Moreno-Iribas C, Engstrom G, Sonestedt E, Lind M, Otten J, Khaw KT, Aune D, Riboli E, Wareham NJ, Imamura F, Forouhi NG, di Angelantonio E, Wood AM, Butterworth AS, Perez-Cornago A. The associations of major foods and fibre with risks of ischaemic and haemorrhagic stroke: a prospective study of 418 329 participants in the EPIC cohort across nine European countries. Eur Heart J 2020.

3. Davey Smith G, Hemani G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. Hum Mol Genet 2014;**23**(R1):R89-98.

4. Itan Y, Jones BL, Ingram CJ, Swallow DM, Thomas MG. A worldwide correlation of lactase persistence phenotype and genotypes. BMC Evol Biol 2010;**10**:36.

5. Enattah NS, Sahi T, Savilahti E, Terwilliger JD, Peltonen L, Jarvela I. Identification of a variant associated with adult-type hypolactasia. Nat Genet 2002;**30**(2):233-7.

6. Torniainen S, Hedelin M, Autio V, Rasinpera H, Balter KA, Klint A, Bellocco R, Wiklund F, Stattin P, Ikonen T, Tammela TL, Schleutker J, Gronberg H, Jarvela I. Lactase persistence, dietary intake of milk, and the risk for prostate cancer in Sweden and Finland. Cancer Epidemiol Biomarkers Prev 2007;**16**(5):956-61.

7. Smith GD, Lawlor DA, Timpson NJ, Baban J, Kiessling M, Day IN, Ebrahim S. Lactase persistence-related genetic variant: population substructure and health outcomes. Eur J Hum Genet 2009;**17**(3):357-67.

8. Travis RC, Appleby PN, Siddiq A, Allen NE, Kaaks R, Canzian F, Feller S, Tjonneland A, Fons Johnsen N, Overvad K, Ramon Quiros J, Gonzalez CA, Sanchez MJ, Larranaga N, Chirlaque MD, Barricarte A, Khaw KT, Wareham N, Trichopoulou A, Valanou E, Oustoglou E, Palli D, Sieri S, Tumino R, Sacerdote C, Bueno-de-Mesquita HB, Stattin P, Ferrari P, Johansson M, Norat T, Riboli E, Key TJ. Genetic variation in the lactase gene, dairy product intake and risk for prostate cancer in the European prospective investigation into cancer and nutrition. Int J Cancer 2013;**132**(8):1901-10.

9. Lamri A, Poli A, Emery N, Bellili N, Velho G, Lantieri O, Balkau B, Marre M, Fumeron F. The lactase persistence genotype is associated with body mass index and dairy consumption in the D.E.S.I.R. study. Metabolism 2013;**62**(9):1323-9.

10. Vissers LET, Sluijs I, van der Schouw YT, Forouhi NG, Imamura F, Burgess S, Barricarte A, Boeing H, Bonet C, Chirlaque MD, Fagherazzi G, Franks PW, Freisling H, Gunter MJ, Quiros JR, Ibsen DB, Kaaks R, Key T, Khaw KT, Kuhn T, Mokoroa O, Nilsson PM, Overvad K, Pala V, Palli D, Panico S, Sacerdote C, Spijkerman AMW, Tjonneland A, Tumino R, Rodriguez-Barranco M, Rolandsson O, Riboli E, Sharp SJ, Langenberg C, Wareham NJ. Dairy Product Intake and Risk of Type 2 Diabetes in EPIC-InterAct: A Mendelian Randomization Study. Diabetes Care 2019;**42**(4):568-575.

11. Bergholdt HK, Nordestgaard BG, Varbo A, Ellervik C. Milk intake is not associated with ischaemic heart disease in observational or Mendelian randomization analyses in 98,529 Danish adults. Int J Epidemiol 2015;**44**(2):587-603.

12. Ding M, Huang T, Bergholdt HK, Nordestgaard BG, Ellervik C, Qi L. Dairy consumption, systolic blood pressure, and risk of hypertension: Mendelian randomization study. Bmj 2017;**356**:j1000.

13. Yang Q, Lin SL, Au Yeung SL, Kwok MK, Xu L, Leung GM, Schooling CM. Genetically predicted milk consumption and bone health, ischemic heart disease and type 2 diabetes: a Mendelian randomization study. Eur J Clin Nutr 2017.

14. Smith CE, Coltell O, Sorli JV, Estruch R, Martinez-Gonzalez MA, Salas-Salvado J, Fito M, Aros F, Dashti HS, Lai CQ, Miro L, Serra-Majem L, Gomez-Gracia E, Fiol M, Ros E, Aslibekyan S, Hidalgo B, Neuhouser ML, Di C, Tucker KL, Arnett DK, Ordovas JM, Corella D. Associations of the MCM6-rs3754686 proxy for milk intake in Mediterranean and American populations with cardiovascular biomarkers, disease and mortality: Mendelian randomization. Sci Rep 2016;**6**:33188.

15. Riboli E, Hunt KJ, Slimani N, Ferrari P, Norat T, Fahey M, Charrondiere UR, Hemon B, Casagrande C, Vignat J, Overvad K, Tjonneland A, Clavel-Chapelon F, Thiebaut A, Wahrendorf J, Boeing H, Trichopoulos D, Trichopoulou A, Vineis P, Palli D, Bueno-De-Mesquita HB, Peeters PH, Lund E, Engeset D, Gonzalez CA, Barricarte A, Berglund G, Hallmans G, Day NE, Key TJ, Kaaks R, Saracci R. European Prospective Investigation into Cancer and Nutrition (EPIC): study populations and data collection. Public Health Nutr 2002;**5**(6b):1113-24.

16. Langenberg C, Sharp S, Forouhi NG, Franks PW, Schulze MB, Kerrison N, Ekelund U, Barroso I, Panico S, Tormo MJ, Spranger J, Griffin S, van der Schouw YT, Amiano P, Ardanaz E, Arriola L, Balkau B, Barricarte A, Beulens JW, Boeing H, Bueno-de-Mesquita HB, Buijsse B,

Chirlaque Lopez MD, Clavel-Chapelon F, Crowe FL, de Lauzon-Guillan B, Deloukas P, Dorronsoro M, Drogan D, Froguel P, Gonzalez C, Grioni S, Groop L, Groves C, Hainaut P, Halkjaer J, Hallmans G, Hansen T, Huerta Castano JM, Kaaks R, Key TJ, Khaw KT, Koulman A, Mattiello A, Navarro C, Nilsson P, Norat T, Overvad K, Palla L, Palli D, Pedersen O, Peeters PH, Quiros JR, Ramachandran A, Rodriguez-Suarez L, Rolandsson O, Romaguera D, Romieu I, Sacerdote C, Sanchez MJ, Sandbaek A, Slimani N, Sluijs I, Spijkerman AM, Teucher B, Tjonneland A, Tumino R, van der AD, Verschuren WM, Tuomilehto J, Feskens E, McCarthy M, Riboli E, Wareham NJ. Design and cohort description of the InterAct Project: an examination of the interaction of genetic and lifestyle factors on the incidence of type 2 diabetes in the EPIC Study. Diabetologia 2011;**54**(9):2272-82.

17. Danesh J, Saracci R, Berglund G, Feskens E, Overvad K, Panico S, Thompson S, Fournier A, Clavel-Chapelon F, Canonico M, Kaaks R, Linseisen J, Boeing H, Pischon T, Weikert C, Olsen A, Tjonneland A, Johnsen SP, Jensen MK, Quiros JR, Svatetz CA, Perez MJ, Larranaga N, Sanchez CN, Iribas CM, Bingham S, Khaw KT, Wareham N, Key T, Roddam A, Trichopoulou A, Benetou V, Trichopoulos D, Masala G, Sieri S, Tumino R, Sacerdote C, Mattiello A, Verschuren WM, Bueno-de-Mesquita HB, Grobbee DE, van der Schouw YT, Melander O, Hallmans G, Wennberg P, Lund E, Kumle M, Skeie G, Ferrari P, Slimani N, Norat T, Riboli E. EPIC-Heart: the cardiovascular component of a prospective study of nutritional, lifestyle and biological factors in 520,000 middle-aged participants from 10 European countries. Eur J Epidemiol 2007;**22**(2):129-41.

18. Kroke A, Klipstein-Grobusch K, Voss S, Moseneder J, Thielecke F, Noack R, Boeing H. Validation of a self-administered food-frequency questionnaire administered in the European Prospective Investigation into Cancer and Nutrition (EPIC) Study: comparison of energy, protein, and macronutrient intakes estimated with the doubly labeled water, urinary nitrogen, and repeated 24-h dietary recall methods. Am J Clin Nutr 1999;**70**(4):439-47.

19. Margetts BM, Pietinen P. European Prospective Investigation into Cancer and Nutrition: validity studies on dietary assessment methods. Int J Epidemiol 1997;**26 Suppl 1**:S1-5.

20. Beulens JW, Monninkhof EM, Verschuren WM, van der Schouw YT, Smit J, Ocke MC, Jansen EH, van Dieren S, Grobbee DE, Peeters PH, Bueno-de-Mesquita HB. Cohort profile: the EPIC-NL study. Int J Epidemiol 2010;**39**(5):1170-8.

21. Rodriguez S, Gaunt TR, Day IN. Hardy-Weinberg equilibrium testing of biological ascertainment for Mendelian randomization studies. Am J Epidemiol 2009;**169**(4):505-14.

22. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet 2006;**38**(8):904-9.

23. Dzialanski Z, Barany M, Engfeldt P, Magnuson A, Olsson LA, Nilsson TK. Lactase persistence versus lactose intolerance: Is there an intermediate phenotype? Clin Biochem 2016;**49**(3):248-52.

24. Burgess S, Small DS, Thompson SG. A review of instrumental variable estimators for Mendelian randomization. Stat Methods Med Res 2015.

25. Prentice RL. A case-cohort design for epidemiologic cohort studies and disease prevention trials. Biometrika 1986;**73**(1):1-11.

26. Sahi T. The inheritance of selective adult-type lactose malabsorption. Scand J Gastroenterol Suppl 1974;**30**:1-73.

27. Team RC. *R: A language and environment for statistical computing.* https://www.R-project.org/.

28. Borgan Ø. Modeling Survival Data: Extending the Cox Model. Terry M. Therneau and Patricia M. Grambsch, Springer-Verlag, New York, 2000. ISBN 0-387-98784-3. Stat Med 2001;**20**(13):2053-2054.

29. Viechtbauer W. Conducting Meta-Analyses in R with the metafor Package. 2010 2010;**36**(3):48.

30. M. Gordon TL. forestplot: Advances Forest Plot Using 'grid' Graphics. R package Version 1.7.2. . 2017

31. Holmes MV, Ala-Korpela M, Smith GD. Mendelian randomization in cardiometabolic disease: challenges in evaluating causality. Nat Rev Cardiol 2017.

32. Corgneau M, Scher J, Ritie-Pertusa L, Le DTL, Petit J, Nikolova Y, Banon S, Gaiani C. Recent advances on lactose intolerance: Tolerance thresholds and currently available answers. Crit Rev Food Sci Nutr 2017;**57**(15):3344-3356.

33. Voedingsstoffenbestand SN. NEVO-tabel 2016 (Dutch food composition table). In; 2016.

34. Kettunen J, Silander K, Saarela O, Amin N, Muller M, Timpson N, Surakka I, Ripatti S, Laitinen J, Hartikainen AL, Pouta A, Lahermo P, Anttila V, Mannisto S, Jula A, Virtamo J, Salomaa V, Lehtimaki T, Raitakari O, Gieger C, Wichmann EH, Van Duijn CM, Smith GD, McCarthy MI, Jarvelin MR, Perola M, Peltonen L. European lactase persistence genotype shows evidence of association with increase in body mass index. Hum Mol Genet 2010;**19**(6):1129-36.

35. Key TJ, Appleby PN, Bradbury KE, Sweeting M, Wood A, Johansson I, Kuhn T, Steur M, Weiderpass E, Wennberg M, Lund Wurtz AM, Agudo A, Andersson J, Arriola L, Boeing H, Boer JMA, Bonnet F, Boutron-Ruault MC, Cross AJ, Ericson U, Fagherazzi G, Ferrari P, Gunter M, Huerta JM, Katzke V, Khaw KT, Krogh V, La Vecchia C, Matullo G, Moreno-Iribas C, Naska A,

Nilsson LM, Olsen A, Overvad K, Palli D, Panico S, Molina-Portillo E, Quiros JR, Skeie G, Sluijs I, Sonestedt E, Stepien M, Tjonneland A, Trichopoulou A, Tumino R, Tzoulaki I, van der Schouw YT, Verschuren WMM, di Angelantonio E, Langenberg C, Forouhi N, Wareham N, Butterworth A, Riboli E, Danesh J. Consumption of Meat, Fish, Dairy Products, and Eggs and Risk of Ischemic Heart Disease. Circulation 2019;**139**(25):2835-2845.

36. Meschia JF, Bushnell C, Boden-Albala B, Braun LT, Bravata DM, Chaturvedi S, Creager MA, Eckel RH, Elkind MS, Fornage M, Goldstein LB, Greenberg SM, Horvath SE, Iadecola C, Jauch EC, Moore WS, Wilson JA. Guidelines for the primary prevention of stroke: a statement for healthcare professionals from the American Heart Association/American Stroke Association. Stroke 2014;**45**(12):3754-832.

37. van Mierlo LA, Arends LR, Streppel MT, Zeegers MP, Kok FJ, Grobbee DE, Geleijnse JM. Blood pressure response to calcium supplementation: a meta-analysis of randomized controlled trials. J Hum Hypertens 2006;**20**(8):571-80.

38. Geleijnse JM, Kok FJ, Grobbee DE. Blood pressure response to changes in sodium and potassium intake: a metaregression analysis of randomised trials. J Hum Hypertens 2003;**17**(7):471-80.

39. Cordain L, Eaton SB, Sebastian A, Mann N, Lindeberg S, Watkins BA, O'Keefe JH, Brand-Miller J. Origins and evolution of the Western diet: health implications for the 21st century. Am J Clin Nutr 2005;**81**(2):341-54.

Table 1. Habitual dietary intake and association between lactase persistence genotype and dietary intake among EPIC-CVD subcohort participants.

	Dietary intake	β*	95% CI		p-value†	I^2 (%)‡	N
Total energy (kcal/day)	2051 [1661, 2528]	-2.9	-19.9	14.1	0.74	0.0	13,114
Milk (g/day)	165 [43, 302]	13.7	8.4	19.1	$4.9*10^{-7}$	69.3	13,114
Non-milk dairy (g/day)	96 [46, 176]	2.1	-0.8	5.0	0.15	0.0	13,114
Milk for coffee and creamers	0 [0, 3]	1.0	0.0	2.0	0.06	0.0	9,048
Dairy creams	0 [0, 0]	-0.1	-0.3	0.1	0.32	0.0	6,331
Milk based puddings	1 [0, 3]	0.0	-0.1	0.2	0.72	0.0	11,826
Curd	0 [0, 11]	0.2	-0.6	1.0	0.59	0.0	9,272
Yogurt, thick fermented milk	0 [0, 4]	-0.5	-1.1	0.0	0.06	0.0	9,272
Ice cream	25 [0, 97]	0.7	-1.6	3.1	0.54	0.0	13,114
Cheese	3 [0, 9]	-0.1	-0.4	0.2	0.49	0.0	13,114
Vegetables (g/day)	27 [14, 50]	-0.2	-1.0	0.7	0.72	56.0	13,114
Fruit (g/day)	152 [98, 234]	-2.3	-5.2	0.6	0.12	43.5	13,114
Meat and meat products (g/day)	186 [100, 306]	-7.3	-12.1	-2.5	$3.0*10^{-3}$	85.7	13,114
Fish and shellfish (g/day)	103 [69, 142]	-0.9	-2.4	0.6	0.22	57.5	13,114
Soft drinks (g/day)	28 [15, 51]	0.3	-0.5	1.1	0.46	64.6	13,114
Coffee (g/day)	22 [17, 27]	-0.8	-4.8	3.2	0.70	64.8	13,114
Tea (g/day)	7 [0, 81]	6.2	-1.8	14.1	0.13	0.0	13,114
Alcohol (g/day)	288 [91, 580]	-5.9	-11.5	-0.4	0.04	0.0	13,114

Dietary intake data is described as median [p25,p75]. * Beta derived from linear regression model, investigating association between additional rs4988235 T alleles with dietary intake, adjusted for sex, age, two genetic PC, and study centre. † p-value of linear regression model. ‡ I^2 for fixed effect meta-analysis of analyses among full EPIC-CVD subcohort stratified by participants with hard call and imputed rs4988235 genotype.

Table 2. Habitual dietary intake and association between lactase persistence genotype and dietary intake among 2,025 EPIC-NL subcohort participants.

	Dietary intake	β*	95% CI		p-value
Total energy intake (kcal/day)	1911 [1625, 2296]	-17.4	-52.1	17.3	0.32
Milk (g/day)	219 [74, 410]	36.7	19.8	53.6	$2.2*10^{-5}$
Unfermented, unsweetened milk	111 [27, 245]	20.9	7.8	34.1	$1.8*10^{-3}$
Buttermilk	2 [0, 100]	11.3	1.0	21.7	0.03
Sweetened milk	15 [0, 31]	4.4	1.7	7.1	$1.6*10^{-3}$
Non-milk dairy (g/day)	158 [92, 212]	0.6	-8.1	9.2	0.90
Curd	7 [2, 12]	-0.4	-1.3	0.5	0.39
Yogurt	51 [15, 96]	-1.8	-8.3	4.7	0.58
Cheese	30 [20, 47]	-1.1	-3.0	0.7	0.23
Vegetables (g/day)	133 [104, 169]	-2.3	-6.0	1.4	0.22
Fruit (g/day)	249 [145, 365]	-0.1	-11.8	11.6	0.99
Meat (g/day)	103 [63, 136]	-4.4	-8.0	-0.9	0.01
Fish (g/day)	8 [3, 15]	0.3	-0.4	1.0	0.37
Beverages †(g/day)	1463 [1161, 1833]	-31.2	-70.1	7.6	0.11
Alcohol (g/day)	6 [1, 17]	0.1	-1.0	1.1	0.93

Dietary intake data is described as median [p25,p75]. * Beta and 95% CI derived from linear regression model, adjusting for sex, age, two genetic PC, and study centre. † Includes coffee, tea, sugar or artificially sweetened beverages, fruit juice and alcoholic beverages.

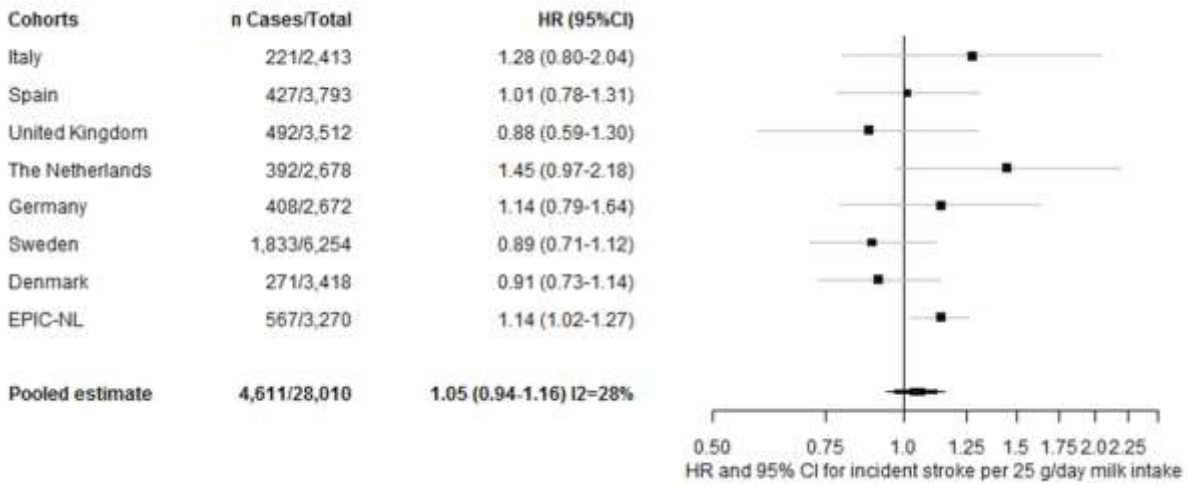

Figure 1. Hazard ratio and 95%CI for each 25 g/day increase in genetically predicted milk intake and risk of total stroke in EPIC-CVD countries and in EPIC-NL, assuming an additive effect of rs4988235.

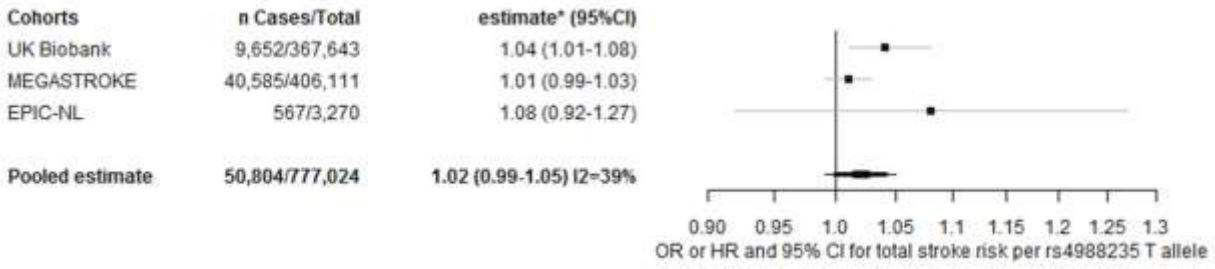

Figure 2. Odds ratio or Hazard ratio and 95%CI for each additional rs4998235 lactase persistence (T) allele and risk of total stroke in UK Biobank, MEGASTROKE (including EPIC-CVD data) and EPIC-NL.

* OR for UK Biobank and MEGASTROKE, HR for EPIC-NL, and RR for the pooled estimate.

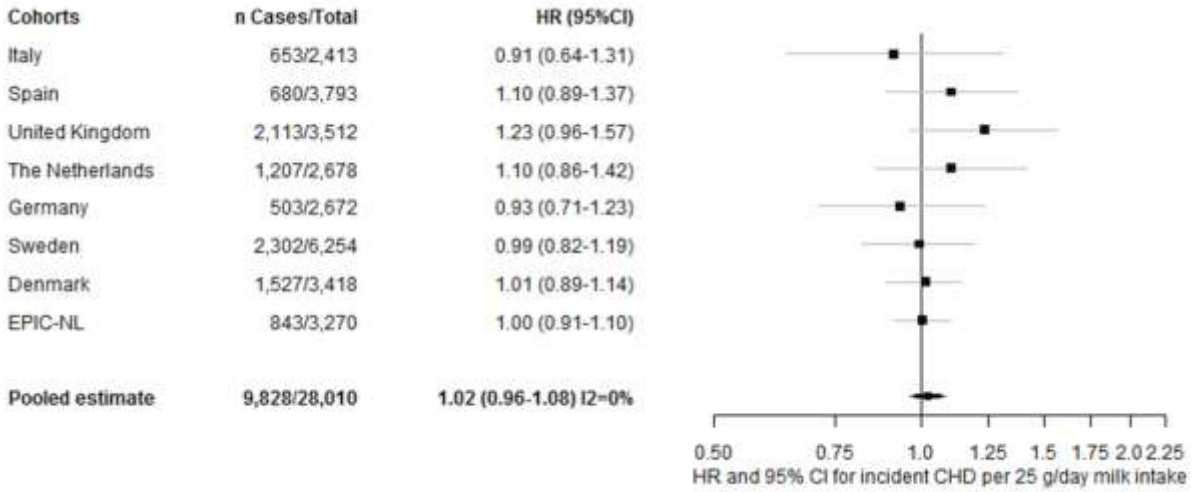

Figure 3. Hazard ratio and 95%CI for each 25 g/day increase in genetically predicted milk intake and risk of CHD in EPIC-CVD countries and in EPIC-NL, assuming an additive effect of rs4988235.

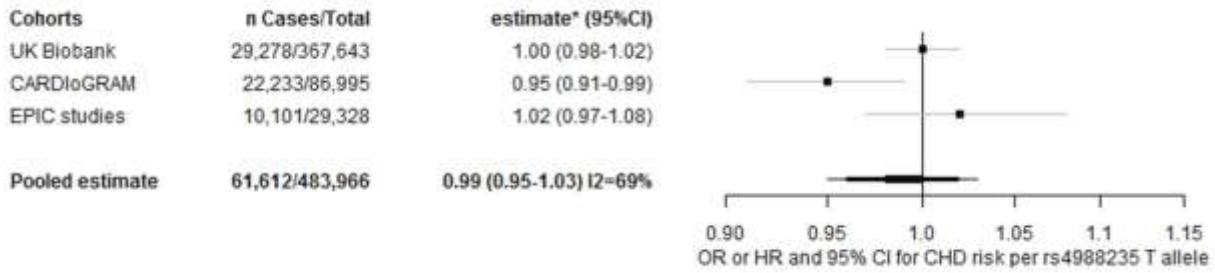

Figure 4. Odds ratio or Hazard ratio and 95%CI for each additional rs4998235 lactase persistence (T) allele and risk of CHD in UK Biobank, CARDIoGRAM and the EPIC studies (EPIC-CVD and EPIC-NL combined).

* OR for UK Biobank and CARDIoGRAM, HR for EPIC-NL, and RR for the pooled estimate.